

Ryan Perry  
IST 632  
Research Paper  
December 10, 2014

### **Digital Special Collections**

At their core, libraries serve two primary functions: collecting and connecting. In collecting information resources, libraries help to preserve the products of knowledge creation; in connecting users to these materials, libraries enable access and implementation of knowledge. Special collections, as storehouses of singular materials, protect an endangered subset of recorded knowledge—without constant stewardship, this knowledge would become extinct. Both preservation and access are important elements to consider in managing a special collection. However, in a harsh climate of scarcity in both time and finances, special collections librarians must prioritize in order to prevent the extinction of significant resources.

Through developments in technology, it has become progressively easier to produce surrogate copies of information resources. Reproducibility serves to increase the divergence between intellectual *content* and the physical *carrier* through which it exists. The aura contained in an item's finer details can contribute significantly to scholarship and understanding of its importance—offering context and nuance not captured by a surrogate copy. On the other hand, intellectual content can convey meaning even when alienated from its carrier. Walter Benjamin, a German philosopher, contends: “the technique of reproduction detaches the reproduced object from the domain of tradition. By making many reproductions it substitutes a plurality of copies for a unique existence. And in permitting the reproduction to meet the beholder or listener in his

own particular situation, it reactivates the object reproduced.”<sup>1</sup> Consequently, both content and carrier offer valuable avenues for assessing the cultural heritage significance and intellectual reach of material in special collections.

In considering digitization’s impact on special collections, the tensions between preservation and access, as well as content and carrier, must be properly weighed. Digital collections cannot eclipse the importance of physical collections in terms of preserving the carriers of cultural heritage. However, in terms of access to intellectual content, digital collections have the potential to reach entirely new audiences. A strong online presence for special collections serves to dramatically expand access to materials, while encouraging researchers to visit the collection in person to experience finer details. As a result, digital special collections have the potential to elevate the stature of a library, by making these materials a public advertisement for the library’s significance.

With this theory as a foundation, special collections librarians can leverage the strengths of digital collections while remaining mindful of the enduring importance of physical collections. Taking the limitation of resources as a starting point, prioritization is paramount to success. Realistic goals can help in reducing backlogs, as can focusing energies on collection strengths. It pays to follow in the footsteps of effective institutions in approaching digital special collection. In addition to research on best practices in the field, I have based the recommendations outlined in this paper on conversations with James DelRosso, Web and Digital Projects Manager at Cornell University’s Catherwood Library, and Michele Combs, Lead

---

<sup>1</sup> Walter Benjamin, *The Work of Art in the Age of Its Mechanical Reproducibility and Other Writings on Media*, eds. Michael William Jennings and Brigid Doherty (Cambridge: Harvard University Press, 2008), 19.

Archivist for Syracuse University's Special Collections Research Center. As practicing stewards of digital collections, DelRosso and Combs offer practical insight into the role and value of special collections in an online environment. My recommendations fall into the following categories: selection of materials, content structure, partnerships, usage data, and finally, the open source sharing of tools.

### **Selection of Materials**

In approaching a collection as a candidate for digitization, the curator must consider the role of surrogate copies in relation to the items of record. Given that the original items should typically remain the focus of preservation efforts, focusing digitization efforts on increasing access to intellectual content has the potential help to streamline the processing of materials. As Ricky Erway and Jennifer Schaffner contend in their summation of OCLC's "Digitization Matters" forum, "scaling up digitization of special collections [...] will compel us to temper our historical emphasis on quality with the recognition that large quantities of digitized special collections materials will better serve our users."<sup>2</sup> Acknowledging that new technologies have required materials to be perennially revisited, the desire to produce one "final" digital surrogate has proven unrealistic. When tape-based media emerged, preservationists assumed that the technology would have staying power, only to have the format threatened by relatively short shelf lives and made obsolete by the compact disc and subsequent technologies. The selecting items for digitization should therefore not rest entirely on preservation as a justification.

---

<sup>2</sup> Ricky Erway and Jennifer Schaffner, "Shifting Gears: Gearing Up to Get Into the Flow," OCLC Programs and Research, 2007, accessed October 6, 2014, <http://www.oclc.org/programs/publications/reports/2007-02.pdf>.

There are, of course, notable exceptions wherein the physical container cannot be relied on to last through another technology cycle. Examples of this category include materials in a state of rapid decay or contained in obsolete formats—such as tape-based media. For items wherein the carrier has an especially short shelf life, all efforts should focus on preserving the content. Ultimately, the item of record should remain the target of preservation efforts, while digital surrogates are used to promote accessibility. Special collections librarians need to do what they can to ensure the longevity of their collections. As such, the role of digital surrogates should be assessed on a collection-by-collection basis.

Taking into account the longevity of the carrier, digital special collections curators should use access as the starting point and work their way back to preservation. Why are libraries collecting and preserving materials if not for their value in being used by researchers? A collection in and of itself is an inanimate body of matter; when put to use it comes alive in conveying meaning. While librarians cannot assume to know what materials will be valuable in the future, beginning with the user's need and developing an online collection to meet the interest of an existing audience will ensure that the collection is used and that digitization efforts are worthwhile.<sup>3</sup>

Research cannot proceed with bits and pieces, a critical mass of materials is necessary to support an argument. Erway and Schaffner describe the ineffectiveness of online collections in this regard:

Early digitization projects have focused on cherry-picking for topical projects. Now we find our vast collections represented by a relatively small number of gorgeous images, lovingly selected, described, and presented in deep web portals. These hand-crafted digital presentations don't begin to reflect the breadth of materials in our institutions.<sup>4</sup>

---

<sup>3</sup> Richard J. Cox, "The Documentation Strategy and Archival Appraisal Principles," 232.

<sup>4</sup> Erway, 5.

Rather than settle for a small number of quality online materials, researchers would be better served by more holistic collections that preserve, to a greater extent, the context afforded by similar items and comparative studies. Being realistic about digital libraries means acknowledging the difficulty in having both quantity of material and quality of description. Special collections libraries should provide access as large a quantity of materials as possible, with a minimal level of description to ensure its findability.

### **Content Structure**

Access to digital information resources is not enough on its own—how digital special collections are structured can be nearly as important as the content itself. Without an effective presentation schema, materials lack the findability to make them useful. In the early stages of online collection development, the mere fact that resources could be made available online was enough to draw attention and garner praise. However, as the Internet has grown in size and scope, the true value lies in the usability of resources.

The Internet Archive ([www.archive.org](http://www.archive.org)) offers an example for how user interface can obstruct access to digital materials. With a collection in excess of fifty petabytes, the Internet Archive has a staggeringly large quantity of materials, spanning a range of media: text, audio, video, software, and web content.<sup>5</sup> The sheer size of the collection, as well as the prevalence of duplicate and near-duplicate materials—contribute to the difficulty in finding and identifying

---

<sup>5</sup> “Frequently Asked Questions,” Internet Archive, accessed December 8, 2014, <https://archive.org/about/faqs.php>. One petabyte is equivalent to 2,000 years of recorded sound (if encoded as an mp3 file). 50 Petabytes, if it contained only plain text documents, could store the entirety of written works in all languages since the dawn of recorded history. Jesus Diaz, “How Large is a Petabyte,” *Gizmodo*, July 8, 2009, accessed December 8, 2014, <http://gizmodo.com/5309889/how-large-is-a-petabyte>.

appropriate materials in the archive. Despite these factors, the impenetrability of the archive ultimately stems from its ineffective user interface. Search results do not display the most relevant information; the lack of effective filters and cross-searches hinder users' attempts at parsing the results. Furthermore, items' details pages fail to provide adequate context to foster linkages to the rest of the collection. In essence, each information resource exists in siloed form, effectively cut-off from the rest of the collection. As valuable as these resources may be, they would be much more significant if the Internet Archive were able to better preserve aspects of collection-level context. The Internet Archive has announced a website redesign, however the new layout amounts to little more than a fresh coat of paint—the underlying issues of findability remain unresolved.

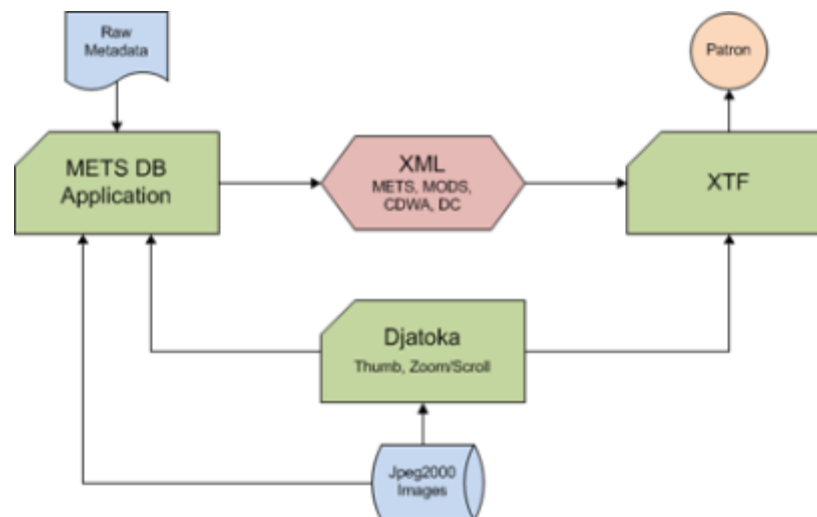
When asked for an example of a model institution for effective digital collections, Michele Combs did not hesitate in her praise of the Online Archive of California (OAC).<sup>6</sup> If the Internet Archive represents a large, but impenetrable collection, the OAC demonstrates how substantial collections need not sacrifice accessibility on account of their size. Similar to the Internet Archive, the OAC represents a collection of collections, in this case integrating the collections of institutions throughout the state. OAC must therefore integrate various organizational standards into a unified whole. Whereas the Internet Archive uses individual items as the distinctive units of the collection, OAC relies on finding aids, which serve to maintain context. A red eye icon denotes which items are available online, and by corollary, which items are only available in the physical collection. This allows users to get a sense for the

---

<sup>6</sup> Michele Combs (Lead Archivist, Syracuse University Special Collections Research Center) in discussion with the author, November 19, 2014.

item's relationship with the rest of the collection, while encouraging users to visit a collection in person to view other items that are unavailable online.

OAC's content structure has as its basis the XTF delivery platform developed by the California Digital Library, a division of the University of California. XTF, or eXtensible Text Framework, is a delivery platform that allows for the discoverability and display of media resources.<sup>7</sup> Syracuse University's Marcel Breuer Digital Archive also uses XTF to deliver its content to users, as outlined in Figure 1 below. Through XTF, media files are displayed via the Djakarta image server, while XML metadata feeds in through a METS (Metadata Encoding and Transmission Standard) database. As an XML schema, METS allows for a mixture of metadata from various content standards (including Dublin Core, MODS, and CDWA)—allowing for flexibility in source material content and format. By structuring content in this manner, it becomes possible to offer dynamic linkages, allowing users to browse through the collection through multiple access points.



**Figure 1:** Content Delivery Framework for Marcel Breuer Digital Archive (<http://breuer.syr.edu/page-about-specs.php>)

<sup>7</sup> "About," eXtensible Text Framework, accessed December 9, 2014, <http://xtf.cdlib.org/about/>.

In browsing the Marcel Breuer Digital Archive, users are able to click through hyperlinked metadata elements to find similar items. For example, clicking on the date of a particular architectural drawing leads to a list of other items in the archive associated with that same date. This could allow for researchers to conduct a comparison of Breuer's style over time, with the ability to draw conclusions on his evolution as an architect. If Syracuse University were to follow the model set by OAC in including links within the finding aids associated with the Breuer Collection, they could further enhance researchers' sense of context in the collection. Currently, however, the finding aid for the Marcel Breuer Papers appear distinct from the digital collection, on the Special Collections Research Center website—with only a single link connecting to the Marcel Breuer Digital Archive as a whole.<sup>8</sup> Users have to cross-check the digital materials with the extensive finding aid in order to determine an item's availability. The power of linked data could have powerful implications for the type of research that would be possible with more collections implementing a dynamic structure in the model of OAC.

### **Building Partnerships**

Even with a well-curated collection and an effective content management system, a digital special collection cannot exist in an online silo. Libraries should both solicit direct partnerships with collections containing similar materials, and also consider their indirect relationships with search engines. Both avenues for outreach can help to build the collection and the institution's reputation.

---

<sup>8</sup> "Marcel Breuer Papers," Syracuse University Libraries, accessed December 9, 2014, [http://library.syr.edu/digital/guides/b/breuer\\_m.htm](http://library.syr.edu/digital/guides/b/breuer_m.htm).



One key element of the Online Archive of California's success lies in the partnerships it has formed with other institutions. In aggregating collections with compatible subject strengths, digital special collections librarians can build user bases who rely on the resources available for multiple projects. With the breadth and depth of California-related resources, the OAC website becomes a destination for users. Furthermore, these partnerships offer linkages from component institutions, whose collections and websites lead users into the larger repository. As Combs notes, items from equivalent institution-level collections have the potential to be aggregated into a unified collection. For example, the OAC has numerous collections relating to the San Francisco earthquake of 1906—a union finding aid incorporating digital items from the twenty-seven related collections listed could add immense value for researchers. Such endeavors serve to leverage the advantages of the online environment to produce tools that would not be possible with purely physical collections.

The Marcel Breuer Digital Archive, attains the ideal of a union collection on a smaller scale and with a narrowly bounded subject. While Syracuse University has led the initiative in creating the archive, the collection includes materials from: Smithsonian Archives of American Art, Bauhaus Archive Berlin, Stiftung Bauhaus Dessau, Harvard University, GTA Archive of Eidgenössische Technische Hochschule Zürich, Vitra Design Museum, and the University of East Anglia. With this diversity of contributing institutions cooperating in producing the digital collection, Syracuse University has produced a multi-faceted collection that can easily function as a coherent whole.

Indirect institutional relationships, such as with search engines like Google and content providers like Wikipedia, can help to drive large amounts of traffic to digital special collections.

Search Engine Optimization (SEO) stands as one of the important advantages to using XTF and METS as tools for content management. Although she did not have the relevant web analytics statistics available, Combs anecdotically vouched for the large quantity of visitors driven to Syracuse University's digital collections via Google and other search engines. A link on Marcel Breuer's Wikipedia entry drives much of the external access to the digital collection. With peak traffic from Wikipedia coinciding with the end of the semesters, this implies that students use the collection in completing their school work.<sup>9</sup> Being mindful of these external linkages can help to drive traffic and attention to digital collections.

### **Usage Assessment**

Once established, digital special collections should be the subject of continuous self-assessment. As the user's needs factor heavily into collection management, usage assessment can produce insights that can be used to better serve a collection's audience. Usage data can provide quantitative figures measuring the collection's impact, which can be used when requesting institutional funding or applying for grants. In addition to usage data related to SEO and user origin, assessment should relate to how users interact with the collection.

Cornell University's DigitalCommons@ILR, while ostensibly serving the School of Industrial and Labor Relations, touts its global reach through a map on its landing page, which charts real-time readership of materials in the collection.<sup>10</sup> This graphic allows visitors to the site to gain a sense for who else is currently using the collection and which items they have selected

---

<sup>9</sup> Combs interview.

<sup>10</sup> In our conversation, James DelRosso indicated that this map is a feature of their content display software. James DelRosso (Web and Digital Projects Manager at Cornell University's Catherwood Library) in discussion with the author, October 6, 2014.

for download. Additionally, the graphic advertises the total number of papers available and the number of downloads across various time periods. While most websites reserve this information for staff use only, Cornell's decision to make it public and prominently displayed has important ramifications. The display of the global impact of the collection serves to reinforce users' perception of the resource as worthy of attention as a globally significant resource—by extension elevating the prominence of the school within the field of labor relations, and Cornell as a research university.

Beyond these surface-level indicators of use, qualitative data can also be gathered through the use of surveys. James DelRosso indicated that the DigitalCommons@ILR had incorporated suggestions from multi-layered user survey data collected over its first year of existence in making adjustments to the website. In conducting a user needs assessment for the California Digital Library, Felicia Poe, offers some valuable conclusions that indicate the nuanced conclusions that can be drawn from this data. Based on her assessments, Poe concludes that users: “welcome the ability to work within a single interface to search several resources; prefer a basic search interface with choice of fielded searching and limits; [feel as though] table of contents and bibliographies are highly valued discovery tools; want database recommendations; [consider] compatibility with EndNote [as] essential.”<sup>11</sup> Armed with this information, managers of digital special collections can investigate new features or address current design limitations on their web interface.

---

<sup>11</sup> Felicia Poe, “Listening to Users: How User Communities Can Inform Design,” DLF Spring Forum, 2006, accessed December 10, 2014, [http://www.cdlib.org/services/uxdesign/docs/DLF\\_Spring2006\\_poe.pdf](http://www.cdlib.org/services/uxdesign/docs/DLF_Spring2006_poe.pdf).

## Open Source

Building on the successes of other institutions can help managers of digital special collections to accelerate the development of a robust and effective online presence. Open source solutions enable libraries to iterate on existing systems rather than building their own or subscribing to expensive services. XTF offers a free, open source alternative to subscription content management systems like CONTENTdm. However, the lack of upfront cost comes at a significant trade-off: XTF must be tailored, implemented, and maintained by in-house staff to fit the needs of an institution, whereas CONTENTdm subscribers can rely on OCLC for technical support and troubleshooting. As Combs notes, not all institutions can justify the additional staffing required to support open source software; it may be cheaper in the long run to go with a subscription service.<sup>12</sup>

Sharing between digital special collections need not be limited to open source software; institutions can stimulate innovation in the field by publicizing their methodologies. Through sharing best practices, institutions also stand to gain from increased exposure and status—which can help to encourage donations and other forms of support. Sharing results can also be a stipulation of grant funding, especially from the National Endowment for the Humanities.

The philosophy of open source raises some further questions concerning the ownership of intellectual content related to the structure of digital special collections. Toward the end of our conversation, Michele Combs raised the question, “who owns metadata?”<sup>13</sup> While libraries spend large amounts of resources in creating descriptions for special collections materials, this

---

<sup>12</sup> Combs interview.

<sup>13</sup> Ibid.

content should not be locked down and protected, but rather should be given freely as a service to the intellectual community.

## **Conclusion**

If libraries generally focus on collecting and connecting, managers of digital special collections must pay particular attention to how their actions will impact the user in these terms. Selection of materials for inclusion should hinge on access more than preservation. Content structures should emphasize the ability of the user to browse naturally between items. Partnerships with other institutions should add value through cooperative finding aids and cross-organizational collections. Usage assessments should drive web design and be used to promote the stature of the collection. Finally, open source solutions should be pursued where possible to contribute to the further development of innovative methods. In working as a cohesive coordinated effort, digital special collections can expand their reach; libraries can feed into a greater appreciation and use of cultural heritage materials. To put it colloquially, the rising tide of cultural heritage research could lift all collections further into the spotlight.

## References

“About.” eXtensible Text Framework. Accessed December 9, 2014. <http://xtf.cdlib.org/about/>.

Benjamin, Walter, *The Work of Art in the Age of Its Mechanical Reproducibility and Other Writings on Media*. Eds. Michael William Jennings and Brigid Doherty. Cambridge: Harvard University Press, 2008.

Combs, Michele (Lead Archivist, Syracuse University Special Collections Research Center) in discussion with the author. November 19, 2014.

Cox, Richard J. "The Documentation Strategy and Archival Appraisal Principles: A Different Perspective." *Archivaria* 1, no. 38 (1994).

DelRosso, James (Web and Digital Projects Manager at Cornell University’s Catherwood Library) in discussion with the author, October 6, 2014.

Diaz, Jesus. “How Large is a Petabyte,” *Gizmodo*. July 8, 2009. Accessed December 8, 2014. <http://gizmodo.com/5309889/how-large-is-a-petabyte>.

Erway, Ricky, and Jennifer Schaffner. “Shifting Gears: Gearing Up to Get Into the Flow.” OCLC Programs and Research. 2007. Accessed October 6, 2014. <http://www.oclc.org/programs/publications/reports/2007-02.pdf>.

“Frequently Asked Questions.” Internet Archive. Accessed December 8, 2014. <https://archive.org/about/faqs.php>.

“Marcel Breuer Papers.” Syracuse University Libraries. Accessed December 9, 2014. [http://library.syr.edu/digital/guides/b/breuer\\_m.htm](http://library.syr.edu/digital/guides/b/breuer_m.htm).

Poe, Felicia. “Listening to Users: How User Communities Can Inform Design.” DLF Spring Forum, 2006. Accessed December 10, 2014/ [http://www.cdlib.org/services/uxdesign/docs/DLF\\_Spring2006\\_poe.pdf](http://www.cdlib.org/services/uxdesign/docs/DLF_Spring2006_poe.pdf).